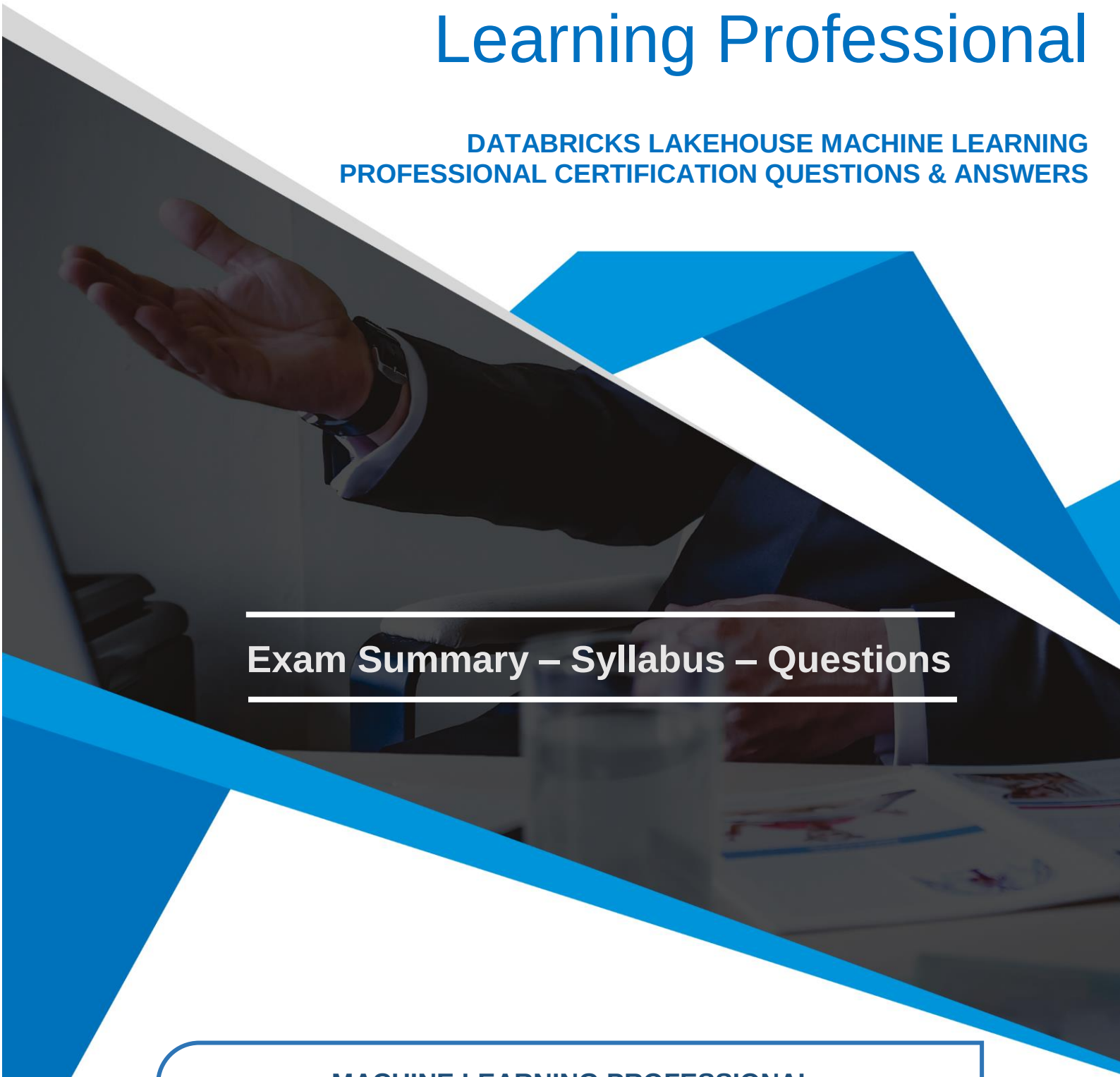
Databricks Machine Learning Professional

DATABRICKS LAKEHOUSE MACHINE LEARNING PROFESSIONAL CERTIFICATION QUESTIONS & ANSWERS



Exam Summary – Syllabus – Questions

MACHINE LEARNING PROFESSIONAL

Databricks Certified Machine Learning Professional

60 Questions Exam – 70% Cut Score – Duration of 120 minutes

www.CertFun.com

Table of Contents

Know Your Machine Learning Professional Certification Well:

The Machine Learning Professional is best suitable for candidates who want to gain knowledge in the Databricks ML Engineer. Before you start your Machine Learning Professional preparation you may struggle to get all the crucial Lakehouse Machine Learning Professional materials like Machine Learning Professional syllabus, sample questions, study guide.

But don't worry the Machine Learning Professional PDF is here to help you prepare in a stress free manner.

The PDF is a combination of all your queries like-

- What is in the Machine Learning Professional syllabus?
- How many questions are there in the Machine Learning Professional exam?
- Which Practice test would help me to pass the Machine Learning Professional exam at the first attempt?

Passing the Machine Learning Professional exam makes you Databricks Certified Machine Learning Professional. Having the Lakehouse Machine Learning Professional certification opens multiple opportunities for you. You can grab a new job, get a higher salary or simply get recognition within your current organization.

Databricks Lakehouse Machine Learning Professional Certification Details:

Exam Name	Databricks Certified Machine Learning Professional
Exam Code	Machine Learning Professional
Exam Price	$200 (USD)
Duration	120 mins
Number of Questions	60
Passing Score	70%
Books / Training	**Machine Learning in Production**
Schedule Exam	**Kryterion Webassesor**
Sample Questions	**Databricks Machine Learning Professional Sample Questions**
Practice Exam	**Databricks Machine Learning Professional Certification Practice Exam**

Machine Learning Professional Syllabus:

Topic	Details	Weights
Experimentation	- Data Management • Read and write a Delta table • View Delta table history and load a previous version of a Delta table • Create, overwrite, merge, and read Feature Store tables in machine learning workflows - Experiment Tracking • Manually log parameters, models, and evaluation metrics using MLflow • Programmatically access and use data, metadata, and models from MLflow experiments - Advanced Experiment Tracking • Perform MLflow experiment tracking workflows using model signatures and input examples • Identify the requirements for tracking nested runs • Describe the process of enabling autologging, including with the use of Hyperopt • Log and view artifacts like SHAP plots, custom visualizations, feature data, images, and metadata	**30%**
Model Lifecycle Management	- Preprocessing Logic • Describe an MLflow flavor and the benefits of using MLflow flavors • Describe the advantages of using the pyfunc MLflow flavor • Describe the process and benefits of including	**30%**

Topic	Details	Weights
	preprocessing logic and context in custom model classes and objects - Model Management • Describe the basic purpose and user interactions with Model Registry • Programmatically register a new model or new model version. • Add metadata to a registered model and a registered model version • Identify, compare, and contrast the available model stages • Transition, archive, and delete model versions - Model Lifecycle Automation • Identify the role of automated testing in ML CI/CD pipelines • Describe how to automate the model lifecycle using Model Registry Webhooks and Databricks Jobs • Identify advantages of using Job clusters over all-purpose clusters • Describe how to create a Job that triggers when a model transitions between stages, given a scenario • Describe how to connect a Webhook with a Job • Identify which code block will trigger a shown webhook • Identify a use case for HTTP webhooks and where the Webhook URL needs to come. • Describe how to list all webhooks and how to delete a webhook	
Model Deployment	- Batch	**25%**

Topic	Details	Weights
	• Describe batch deployment as the appropriate use case for the vast majority of deployment use cases • Identify how batch deployment computes predictions and saves them somewhere for later use • Identify live serving benefits of querying precomputed batch predictions • Identify less performant data storage as a solution for other use cases • Load registered models with load_model • Deploy a single-node model in parallel using spark_udf • Identify z-ordering as a solution for reducing the amount of time to read predictions from a table • Identify partitioning on a common column to speed up querying • Describe the practical benefits of using the score_batch operation - Streaming • Describe Structured Streaming as a common processing tool for ETL pipelines • Identify structured streaming as a continuous inference solution on incoming data • Describe why complex business logic must be handled in streaming deployments • Identify that data can arrive out-of-order with structured streaming • Identify continuous predictions in time-based prediction store as a scenario for streaming	

Topic	Details	Weights
	deployments • Convert a batch deployment pipeline inference to a streaming deployment pipeline • Convert a batch deployment pipeline writing to a streaming deployment pipeline - Real-time • Describe the benefits of using real-time inference for a small number of records or when fast prediction computations are needed • Identify JIT feature values as a need for real-time deployment • Query a Model Serving enabled model in the Production stage and Staging stage • Identify how cloud-provided RESTful services in containers is the best solution for production-grade real-time deployments	
Solution and Data Monitoring	- Drift Types • Compare and contrast label drift and feature drift • Identify scenarios in which feature drift and/or label drift are likely to occur • Describe concept drift and its impact on model efficacy - Drift Tests and Monitoring • Describe summary statistic monitoring as a simple solution for numeric feature drift • Describe mode, unique values, and missing values as simple solutions for categorical feature drift • Describe tests as more robust monitoring solutions for numeric feature drift than simple	**15%**

Topic	Details	Weights
	summary statistics • Describe tests as more robust monitoring solutions for categorical feature drift than simple summary statistics • Compare and contrast Jenson-Shannon divergence and Kolmogorov-Smirnov tests for numerical drift detection • Identify a scenario in which a chi-square test would be useful - Comprehensive Drift Solutions • Describe a common workflow for measuring concept drift and feature drift • Identify when retraining and deploying an updated model is a probable solution to drift • Test whether the updated model performs better on the more recent data	

Databricks Machine Learning Professional Sample Questions:

Question: 1

A data scientist has developed a model to predict whether or not it will rain using the expected temperature and expected cloud coverage.

However, the relationship between expected temperature and whether or not it rains has changed dramatically since the time period of the data on which the model was trained.

Which type of drift is present in the above scenario?

a) Label drift
b) Feature drift
c) Concept drift
d) Prediction drift
e) None of these

Answer: c

Question: 2

Why is it advantageous to include context in custom model classes and objects?

a) It allows for better version control of models
b) It facilitates easier integration with production environments
c) It improves the interpretability of model predictions
d) It ensures that all necessary preprocessing steps are applied consistently

Answer: b, d

Question: 3

A machine learning engineer has developed a decision tree model using scikit-learn, logged the model using MLflow as decision_tree_model, and stored its run ID in the run_id Python variable. They now want to deploy that model by performing batch inference on a Spark DataFrame spark_df.

Which of the following code blocks can they use to create a function called predict that they can use to complete the task?

a)
```
predict = spark.spark_udf(
f"runs:/{run_id}/decision_tree_model"
)
```
b)
```
predict = mlflow.pyfunc.spark_udf(
spark_df,
f"runs:/{run_id}/decision_tree_model"
)
```
c)
```
predict = sklearn.spark_udf(
spark_df,
f"runs:/{run_id}/decision_tree_model"
)
```
d)
```
predict = mlflow.pyfunc.spark_udf(
spark,
f"runs:/{run_id}/decision_tree_model"
)
```

Answer: d

Question: 4

Which operations can be performed on a Feature Store table in MLflow?

a) Create
b) Overwrite
c) Merge
d) Delete

Answer: a, b, c

Question: 5

For Advanced Experiment Tracking, which of the following are true about model signatures in MLflow?

a) They are optional but recommended
b) They define the input and output schema of the model
c) They must be manually updated with each experiment run
d) They are automatically generated for all models

Answer: a, b

Question: 6

A machine learning engineer wants to move their model version model_version for the MLflow Model Registry model model from the None stage to the Staging stage using MLflow Client client.

Which of the following code blocks can they use to accomplish the task?

a)
```
client.transition_model_version_stage(
name=model,
version=model_version,
stage="Staging"
)
```
b)
```
client.transition_model_version_stage(
name=model,
version=model_version,
stage="None"
)
```
c)
```
client.transition_model_stage(
name=model,
version=model_version,
stage="Staging"
)
```
d)
```
client.transition_model__stage(
name=model,
version=model_version,
from="None",
to="Staging"
)
```
e)
```
client.transition_model_version_stage(
name=model,
version=model_version,
from="None",
to="Staging"
)
```

Answer: a

Question: 7

A data scientist has made the suggestion that their team starts using Feature Store in Databricks Machine Learning. The data scientist claims that using Feature Store will meet a number of their feature management needs.

Which of the following will the team need to implement because it is not automatically provided by Feature Store?

a) Share features across workspaces
b) Measure the drift for individual features
c) Discover features used throughout the organization
d) Track where specific feature tables are used
e) Monitor the freshness of feature tables

Answer: b

Question: 8

A data scientist has developed and logged a scikit-learn gradient boosting regressor model model, and then they ended their Spark session and terminated their cluster.

After starting a new cluster, they want to review the estimators_ of the original model object to analyze each of the trees in the trained model.

Which line of code can be used to restore the model object so that estimators_ is available?

a) mlflow.sklearn.load_model(model_uri)
b) client.pyfunc.load_model(model_uri)
c) mlflow.load_model(model_uri)
d) client.list_artifacts(run_id)["estimators.csv"]
e) This can only be viewed in the MLflow Experiments UI

Answer: a

Question: 9

When merging data into a Delta table, which clause are essential to specify the conditions under which a row should be updated or inserted?

a) WHEN MATCHED THEN
b) ON DUPLICATE KEY UPDATE
c) IF EXISTS THEN UPDATE ELSE INSERT
d) ON CONFLICT DO UPDATE

Answer: a, d

Question: 10

Which are types of data drift?

a) Concept drift
b) Model drift
c) Feature drift
d) Data drift

Answer: a, c

Study Guide to Crack Databricks Lakehouse Machine Learning Professional Exam:

- Getting details of the Machine Learning Professional syllabus, is the first step of a study plan. This pdf is going to be of ultimate help. Completion of the syllabus is must to pass the Machine Learning Professional exam.
- Making a schedule is vital. A structured method of preparation leads to success. A candidate must plan his schedule and follow it rigorously to attain success.
- Joining the Databricks provided training for Machine Learning Professional exam could be of much help. If there is specific training for the exam, you can discover it from the link above.
- Read from the Machine Learning Professional sample questions to gain your idea about the actual exam questions. In this PDF useful sample questions are provided to make your exam preparation easy.
- Practicing on Machine Learning Professional practice tests is must. Continuous practice will make you an expert in all syllabus areas.

Reliable Online Practice Test for Machine Learning Professional Certification

Make CertFun.com your best friend during your Databricks Certified Machine Learning Professional exam preparation. We provide authentic practice tests for the Machine Learning Professional exam. Experts design these online practice tests, so we can offer you an exclusive experience of taking the actual Machine Learning Professional exam. We guarantee you 100% success in your first exam attempt if you continue practicing regularly. Don't bother if you don't get 100% marks in initial practice exam attempts. Just utilize the result section to know your strengths and weaknesses and prepare according to that until you get 100% with our practice tests. Our evaluation makes you confident, and you can score high in the Machine Learning Professional exam.

Start Online Practice of Machine Learning Professional Exam by Visiting URL

https://www.certfun.com/databricks/databricks-certified-machine-learning-professional